Webb Nicholas (Orcid ID: 0000-0001-9355-5512)
Chappell Adrian (Orcid ID: 0000-0002-0694-7348)
Courtright Ericha, M (Orcid ID: 0000-0002-2181-4543)
Duniway Michael (Orcid ID: 0000-0002-9643-2785)

Reducing sampling uncertainty in aeolian research to improve change detection

Nicholas P. Webb[1*], Adrian Chappell[2], Brandon L. Edwards[1], Sarah E. McCord[1], Justin W. Van Zee[1], Bradley F. Cooper[1], Ericha M. Courtright[1], Michael C. Duniway[3], Brenton Sharratt[4], Negussie Tedela[5], David Toledo[6]

[1] USDA-ARS Jornada Experimental Range, Las Cruces, NM 88003, USA
[2] School of Earth and Ocean Sciences, Cardiff University, Cardiff, CF10 3AT, Wales, UK
[3] US Geological Survey, Southwest Biological Science Center, Moab, UT 84532, USA
[4] USDA-ARS Northwest Sustainable Agroecosystems Research Unit, Pullman, WA 99164, USA
[5] Bureau of Land Management, San Luis Valley Field Office, Monte Vista, CO 81144, USA
[6] USDA-ARS Northern Great Plains Research Laboratory, Mandan, ND 58554, USA

*Correspondence to:
N.P. Webb, email: nwebb@nmsu.edu

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019JF005042

Abstract: Measurements of aeolian sediment transport support our understanding of mineral dust impacts on Earth and human systems and assessments of aeolian process sensitivities to global environmental change. However, sample design principles are often overlooked in aeolian research. Here, we use high-density field measurements of sediment mass flux across land use and land cover types to examine sample size and power effects on detecting change in aeolian transport. Temporal variances were 1.6 to 10.1 times the magnitude of spatial variances in aeolian transport for six study sites. Differences in transport were detectable for >67% of comparisons among sites using ~27 samples. Failure to detect change with smaller sample sizes suggests that aeolian transport measurements and monitoring are much more uncertain than recognized. We show how small and selective sampling, common in aeolian research, gives the false impression that differences in aeolian transport can be detected, potentially undermining inferences about process and impacting reproducibility of aeolian research.

Plain Language Summary: Aeolian sediment transport, including wind erosion and dust emission, impacts agricultural production and food security, nutrient cycling, water resources, and climate. Measuring aeolian sediment transport is therefore important for developing an understanding of its impacts on Earth systems and society. However, little consideration has been given to how many samples are needed to measure aeolian transport and detect its change across space and through time. We investigate how sample size, design, and decisions about the precision of change detection affect aeolian transport monitoring. Using field measurements, we show that traditional approaches in aeolian research with small sample sizes and selective placement of equipment are often unable to detect change and support robust inferences about aeolian processes. Unless large numbers of samples are used, uncertainty in field measurements can be so large that it undermines our understanding of how and why aeolian sediment transport rates change across space and through time.

Key Points:

- Uncertainties in monitoring limit statistical inferences and our understanding of aeolian process responses to environmental change.
- Small sample sizes lack power to detect change in aeolian transport rates across space and through time.
- Sampling that adequately estimates spatial and temporal variances is needed to monitor aeolian transport with confidence to detect change.

1. Introduction

Wind erosion, sand dune dynamics and dust emission are highly sensitive to the impacts of natural and anthropogenic environmental change (Yizhaq et al., 2009; Arvin et al., 2017; Hooper & Marx, 2018; Kok et al., 2018), but magnitudes of aeolian sediment transport responses to these changes are not well established (Webb & Pierre, 2018). The highly non-linear response of aeolian transport to wind forcing further complicates efforts to quantify its patterns (Durán et al., 2011). Understanding effects of disturbances and environmental change on aeolian processes, and connected feedbacks, requires field measurements and models that are sensitive to biotic and abiotic drivers. Repeated measurements (monitoring) of aeolian transport rates across ecosystems provides a basis for understanding process mechanics, evaluating treatment effects, and parameterizing dust models for broad-scale investigations (e.g., Hoffman et al., 2008; Belnap et al., 2009; Haustein et al., 2015; Nauman et al., 2018). To detect change, unbiased samples should be used to represent the spatial and temporal variability in transport (de Gruijter et al., 2006). However, large uncertainties in monitoring due to spatiotemporal variability in aeolian transport potentially undermine analyses of wind erosion, dune dynamics, and dust cycle sensitivities to environmental change (Sankey et al., 2012; von Holdt et al., 2019).

Large spatial and temporal variability in aeolian transport arises from land surface-atmosphere interactions at different scales (Ellis et al., 2012). Monitoring these dynamics is critical for quantifying aeolian transport patterns and the underlying causes (Sherman et al., 2018). Efforts are being made to measure aeolian transport rates at increasingly high temporal frequencies (e.g., Baas & van den Berg, 2018; Martin et al., 2018) and improve efficiencies of sampling equipment (e.g., Goossens et al., 2000; Goossens & Buck, 2012). These approaches are justified for studies of aeolian process mechanics, but basic principles of sample design that enable assessment of change in transport rates over space and time are frequently ignored. This is evidenced by the large number of aeolian studies that have used small sample sizes (e.g., $n \leq 3$) without appropriate justification, do not report sample sizes, and selectively position equipment within study sites. The consequences are larger uncertainty than recognized in measurements and the false impression that change can be detected (de Gruijter et al., 2006).

Monitoring aeolian transport and its controls requires sample designs that capture effects of boundary-layer interactions across scales over which inferences about aeolian processes are sought. Since aeolian transport rates respond directly to the spatiotemporal variability in wind shear stress over a site (Stout & Zobeck, 1997), monitoring requires sufficient samples to establish unbiased estimates of mean transport rates and their variability over space and time. Sampling at just a few selected locations (small n) frequently in time has been shown to capture some temporal variance but omits spatial variance and inadequately represents total within-site variance in aeolian transport (de Gruijter et al., 2006; Li et al., 2015; Chappell & Baldock, 2016). Unless site-level variance is represented, change detection inference will be statistically invalid and the scientific significance will be more uncertain than recognized (Lenth, 2001).

Here, we use repeated high-density samples of horizontal sediment mass flux across land use and land cover types in the United States to quantify spatiotemporal variability in aeolian transport. We test the effectiveness of stratified random sampling used by the National Wind Erosion Research Network for monitoring horizontal sediment mass flux (Q) and examine effects of sample size and power (the probability that a test will reject a false null hypothesis) on change detection. We demonstrate that small sample sizes ($n < 10$) are often inadequate for monitoring change in Q over space and time. Furthermore, selective placement of Q sampling locations may poorly characterize transport at a site when its spatial variability is large. We show that with planning, a sample design can be established to reduce uncertainty in monitoring Q to enable change detection over time, in response to treatments, and between land use and land cover types.

2. Data and Methods

2.1 Horizontal sediment mass flux data

We acquired spatiotemporal measurements of horizontal sediment mass flux from six National Wind Erosion Research Network sites (Figure 1; Webb et al., 2016). The network uses a standardized sample design, instrumentation, and methods to measure sediment mass fluxes, dust emission, and land surface and meteorological controls over 1.0 ha sites across US agroecological systems. Data were collected at two semi-arid grassland sites with patchy (Holloman site) and homogeneous (Moab site) grass cover, two shrubland sites with large (Jornada site) and small (San Luis Valley site) heterogeneity in shrub canopy height and cover, and two cropland sites that use no tillage (Mandan site) and conventional tillage (Pullman site) crop management (Table 1). Site vegetation structures are typical of those in deserts, rangelands and croplands across dust source regions globally (Webb et al., 2017).

At each site, a stratified random sample design was used to measure the areal horizontal sediment mass flux (including saltating and suspended sediment) using 27 Modified Wilson and Cooke (MWAC) sediment sampler masts, with samplers at 0.10, 0.25, 0.50 and 0.85 m heights (Webb et al., 2015). In the absence of a priori information, the sites were stratified in a regular 3 x 3 square grid (each 33.3 m^2). Within each of the nine grid cells, MWAC masts were located at three random positions; labelled A1, A2, A3 for the first cell through I1, I2, I3 for the ninth cell (Figure 2). Sediment trapped in the MWAC samplers was collected every ~28 days within sampling periods of 6-37 months (Table 1). Sediment was extracted from the samplers using either wet or dry analysis methods and weighed to determine sediment masses following Webb et al. (2015).

We calculated for each MWAC mast and sampling period the vertically-integrated sediment mass fluxes (Q) from the sediment masses normalized by the MWAC inlet areas (2.34×10^{-4} m^2)

using nonlinear least squares regression to fit exponential functions to the data. Following Ellis et al. (2009), we fitted two-parameter or three-parameter functions to the mass flux profiles depending on whether sediment masses of detectable weight (>0.0001 g) were available at three or four heights respectively (96% of total samples). We then integrated from 0 to 1.0 m height and divided by the sampling periods to obtain:

$$Q = \int_0^1 q(z)\, dz, \quad (1)$$

where $q(z)$ is the sediment mass collected per unit inlet area (m^{-2}) per sampling period (day) at heights z (m), and Q is expressed with units of g m^{-1} day^{-1}.

2.2 Analysis of spatial and temporal variances

We first explored the spatial and temporal contributions to variance in aeolian transport. We plotted Q to test for normality and applied a log-transformation to Q prior to statistical analyses. Following Horvitz and Thompson (1952), if a variable (e.g., Q) is monitored over a land cover type using a stratified random sample, then inverse probability weighting should be used to account for different proportions of observations within strata. The Horvitz-Thompson (HT) estimator of the population mean of the site ($\bar{z}_{HT}$) for the variable (z) is defined as:

$$\bar{z}_{HT} = \frac{\sum_{i=1}^{m} \frac{z_i}{\pi_i}}{n}, \quad (2)$$

where π_i is the probability that the ith sample unit is included in the sample, m = 3 is the simple random sample set size within a stratum, n (up to 27) is the total sample size, and z_i is the value at sampling location i (Horvitz & Thomson, 1952). The $\bar{z}_{HT}$ is population unbiased, meaning that repeated sampling, measurement and calculation would find on average the true value for the mean. The unbiased condition remains if the errors are purely random (zero on average). As the strata and sample sizes were the same at all study locations, the selection probabilities (π_i = 1) were the same for the samples and the HT estimator of the mean produced approximately the same means but smaller spatial variances than a simple random sample. Assuming unbiased sampling, we calculated mean ln(Q) across MWAC sample locations for each sampling period at each site, and mean ln(Q) through time for each MWAC sampler location at each site. For each sampling period we then calculated the spatial variance of the HT estimator of the mean ln(Q) across MWAC sampler locations in space at each site as:

$$\widehat{V_s}(\bar{z}_{HT}) = \frac{\sum_{i=1}^{n} \left(\frac{1-\pi_i}{\pi_i}\right) X_i^2 + \sum_{i=1}^{n} \sum_{j \neq 1}^{n} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_i}\right) X_i X_j}{n^2}, \quad (3)$$

where π_{ij} is the inclusion probability of X_i and X_j population units being in the sample size (Horvitz & Thomson, 1952). We established the mean spatial variance in ln(Q) for each site as $\bar{\widehat{V_s}}$. We then calculated the temporal variance of ln(Q) between sampling periods for each MWAC sampler location as:

$$\widehat{V}_t(\bar{z}) = \frac{1}{(n_t-1)} \sum_{i=1}^{n_t} (z_i - \bar{z}_t)^2, \tag{4}$$

where n_t is the number of sampling periods, and $\bar{z}_t$ is the temporal mean of an MWAC sampler, then established the mean temporal variance in ln(Q) for each site as $\hat{\bar{V}}_t$ (following de Gruijter et al., 2006). We calculated the relative magnitude of the temporal and spatial variances in ln(Q) as the ratio $\hat{\bar{V}}_t/\hat{\bar{V}}_s$.

2.3 Establishing detectable change

Having calculated the magnitudes of variances in ln(Q), we established the effect of sample size and power on detectable change in the transport rate. To examine effects of temporal covariance among samplers on change detection, we calculated lag autocorrelation functions (ρ_k) and MDC for a set of lag periods (k = 1 to 12 months) across ln(Q) for the samples at each site following Priestley (1982) (Figure S1). Following Woodward (1992), the one-tailed test (for change with direction) statistic is commonly based on the t-test:

$$(X_{1-\alpha} + X_{1-\beta})^2 = \frac{\hat{\bar{d}}_{2,1}{}^2}{\frac{V(\hat{\bar{z}}(t_1))}{n_1} + \frac{V(\hat{\bar{z}}(t_2))}{n_2}}, \tag{5}$$

where X is a standard normal distribution, α is the size of the significance test, $1 - \beta$ is the power of the test, n is the sample size, t_1 and t_2 denote the two sampling periods, and $\hat{\bar{d}}_{2,1}$ is the mean difference in estimated means. Equation (5) can be rearranged for $\hat{\bar{d}}_{2,1}$ to establish the difference between means that is dependent on the specified power and size of test (de Gruijter et al., 2006). For a one-sided test this gives:

$$\hat{\bar{d}}_{2,1} = \left(X_{1-\alpha} + X_{1-\beta}\right)\left(\frac{(1-\rho_k)V\left(\hat{\bar{z}}(t_1)\right)}{n_1} + \frac{(1-\rho_k)V\left(\hat{\bar{z}}(t_2)\right)}{n_2}\right)^{0.5}, \tag{6}$$

including the autocorrelation term ρ_k at lag k (here 1 month) to moderate the variances for temporal autocorrelation among samplers (Brus & Noij, 2008). We used $\hat{\bar{d}}_{2,1}$to describe the MDC in ln(Q) – the smallest difference between means that could be detected at a chosen confidence level. We applied Equation (6) with α = 0.05 (Type I error – to infer change when there was none) and β = 0.05 and 0.20 (Type II error – to infer no change when there really was) to calculate MDC between each sampling period at the sites using the HT estimators of the spatial variance for the stratified random samples (Equation 3). We calculated absolute differences between mean ln(Q) of each sampling period and their respective MDC. We then calculated the percentage of time over the sampling periods for which differences in mean ln(Q) were greater than MDC. That is, the frequency at which differences in aeolian transport over time could be detected with 80% confidence and 95% confidence that the differences were significant for α = 0.05. We then calculated MDC for a set of site comparisons with α = 0.05 and for β = 0.05 and 0.20 to evaluate the frequency at which differences in aeolian transport could be detected between sites. We used β = 0.05 and 0.20 for these analyses to examine how sample power affects change detection where the risk of falsely detecting change, or not detecting change, has different implications. For example, a higher level of confidence may be desired to understand physical relations governing aeolian transport versus that needed to decide to initiate management actions to mitigate erosion in a crop field. Finally, we calculated the sample sizes that would be required to detect a difference in aeolian transport; for example, due to land use or land cover change. Assuming $n_1 = n_2 = n$, $\alpha = 0.05$ and $\beta = 0.05$ and the spatial variance and temporal autocorrelation ρ_k of aeolian transport remain constant between comparisons, we used a selection of large and small $\hat{V}_s(\bar{z}_{HT})$ from each site to calculate:

$$n = \frac{\left(X_{1-\alpha}+X_{1-\beta}\right)^2(2-2\rho_k)\hat{V}_s(\bar{z}_{HT})}{(\hat{\bar{d}}_{2,1})^2}. \quad (7)$$

3. Results

Monthly Q showed large variability in space and time across the land cover types, spanning at least two orders of magnitude (Table 2). Overall, Q tended to be smallest at the Holloman grassland and Mandan no-till cropland site, and largest at the San Luis Valley shrubland and Pullman conventional tillage cropland site (Table 2). The temporal distribution of Q varied considerably and by different amounts among MWAC sampler locations within each site, and among land cover types, within large variability of mean Q for the sites through time. For example, at the Jornada site some sampler locations consistently measured larger Q, while others consistently measured smaller Q (Figure 3). The spatial variances were consistently large as the number of samples increased over longer analysis periods. Large spatiotemporal variability in aeolian transport has previously been reported from landscapes around the world (e.g., Sterk & Stein, 1997; Chappell et al., 2003).

We found more variance in Q through time than across space (Table 2). The mean temporal variance in sediment transport ($\hat{\bar{V}}_t$) was larger than the mean spatial variance ($\hat{\bar{V}}_s$) by a factor of 1.6 to 3.4 at all sites except the Holloman grassland sites where $\hat{\bar{V}}_t$ was 10.1 times larger than $\hat{\bar{V}}_s$. Despite $\hat{\bar{V}}_t$ being larger in all cases, both the spatial variance and temporal variance of aeolian transport were different among land use and land cover types. Our results suggest that the spatial variance in aeolian transport may be considerably smaller than the temporal variance for land cover types with homogeneous roughness (e.g., grasslands) and larger at sites with heterogeneous roughness (e.g., shrublands) and exposed bare soils.

The effect of spatiotemporal variability in Q on aeolian transport change detection at the National Wind Erosion Research Network sites is demonstrated by the proportion of comparisons for which measured differences in the transport rate exceeded the MDC (Figure 4). Between-site differences in aeolian transport were detected with 95% confidence for 67% to 91% of site-to-site comparisons (Figure 4a). Reducing the sample power ($\beta = 0.20$) increased the frequency at which differences in aeolian transport between sites were detected between Jornada and Holloman and between Moab and Pullman but had no effect on the other between-site comparisons ($\alpha = 0.05$).

Differences in aeolian transport through time were detected with 95% confidence (at $\alpha = 0.05$) for 80% and 86% of sampling periods at the Holloman and Moab grassland sites respectively (Figure 4b). Differences in aeolian transport were detected for 75% and 87% of sampling periods at the Jornada and San Luis Valley shrubland sites respectively, and for 89% and 60% of comparisons at the Mandan and Pullman cropland sites. Reducing the sample power ($\beta = 0.20$) improved the frequency at which differences in aeolian transport through time were detected at the Holloman grassland site and Jornada shrubland site (by 6%), Mandan cropland site (by 11%) and Pullman cropland site (by 20%), but reduced confidence that differences were statistically significant. The temporal pattern of aeolian transport was random at all sites with only weak temporal autocorrelation among samples at $k = 1$ month at the Jornada shrubland site ($\rho_k = 0.4$, $p < 0.05$) (Figure S1). Monitoring over a longer period (e.g., >5 years) may enable a more robust analysis of temporal autocorrelation among samples. The MDC in aeolian transport did not decrease with increasing lag time at any site (Figure S2). That is, there was no appreciable change detection benefit of measuring Q frequently (monthly) or comparing measurements between longer increments of time.

The MDC varied among land cover types and with time within sites but decreased with larger sample sizes. Figure 5 shows the effect of sample size on the minimum detectable change in aeolian transport with 95% confidence that change was significant ($\alpha = 0.05$). The stratified random sample design used by the National Wind Erosion Research Network ($n = 27$) was able to detect statistically significant differences in Q for the majority of site-to-site and month-to-month comparisons (Figure 4). However, if a smaller sample size is used (e.g., $n < 10$), large differences in aeolian transport must be measured to have confidence that differences are statistically significant; that there was a change in aeolian transport through time or between sites or treatments. While three or fewer sediment sampler masts commonly used in aeolian research might provide a reasonable estimate of the temporal variance in Q, depending on sampler location and measurement period (Figure 3), differences in aeolian transport of 20% to 700% or more between sampling periods or treatments may be required for confident change detection.

4. Discussion and Conclusions

Using repeated high-density measurements of horizontal sediment mass flux across land use and land cover types, we tested the effects of sample size and power on aeolian transport monitoring. Our results show that small samples ($n < 10$) are likely to produce large uncertainties and may be ineffective for characterizing transport rates and detecting change between sites and over time. Selective sampling – for example, placing sediment samplers in specific locations relative to vegetation (e.g., open gaps) – may also bias estimates of transport when its variability is not accounted for. The spatial and temporal variances of aeolian transport can be large and both must be measured to enable change detection with confidence, not overestimate the magnitude of treatment effects, and be reproducible. Sample designs that effectively measure $\hat{V}_s$, implemented over long enough periods of time to establish $\hat{V}_t$, are necessary to monitor and detect change in aeolian transport within and among land cover types.

The sample design used by the National Wind Erosion Research Network was generally sufficient for detecting statistically significant change in Q. At the 95% confidence level, up to 91% of comparisons over space could detect change between sites and up to 89% of comparisons through time could detect change within sites. However, detection varied among sites and over time and was smallest at the Jornada shrubland site. Our results suggest that shrubland sites with a heterogeneous distribution of bare soil, grasses, forbs and shrubs may be the most challenging to sample adequately because the spatial variance of Q is large (Gillette et al., 2006; Gonzales et al., 2018). The financial costs of data collection suggest that implementing a sample design that does not enable consistent change detection (e.g., selective, small n designs) is likely a poor use of resources (Chappell et al., 2003; Li et al., 2015). The costs for making robust inferences from field data are very large if under-sampling gives the false impression that change can be detected – potentially undermining new insights about aeolian processes and causing confusion or lack of confidence in scientific advice about the effectiveness of management options (Chappell & Baldock, 2016). Reducing sample power could be used if larger uncertainty in change detection is acceptable (Desu & Raghavarao, 1990). However, our results show that while reducing sample power had a large effect (6-20%) at some sites, it had no effect on detecting change at other sites. Our results also show that a sample size of ~30 would generally still be needed to consistently detect change at the plot scale (1 ha). The statistical risk of choosing a lower confidence level, possibly producing a false conclusion, should be determined by the data application to research, model parameterization, or management (Smith et al., 2014).

If we seek high confidence in the patterns and processes revealed by aeolian transport measurements, it is necessary to plan sampling accordingly. Consistent with Webster & Oliver (1992), our results show that a sample size of ~100 appears necessary to measure the spatial variance of Q and consistently reduce MDC below 100% across land use and land cover types (Figure 5). The requirement for adequate sampling of the variance in Q will hold across spatial scales, meaning that similarly large sample sizes may be needed to detect change in Q at the plot scale (< 10^2 m^2) or regional scale (>10^4 km^2). Sample sizes can be smaller when the spatial variance of Q is small, but our data show that $\hat{V}_s$ is temporally variable and likely difficult to predict. The MDC framework is statistically rigorous and, by allowing for adjustment of sample power, has the flexibility to avoid being overly stringent. The large MDC (>700%) that may result from very small sample sizes (n < 3) suggests that aeolian research implementing small sample sizes and selective sampling is likely to contain inherently greater uncertainty than previously recognized. These uncertainties have inevitably propagated through our understanding of aeolian processes, parameterization of aeolian transport models, and reproducibility of aeolian research.

To encourage rigor in aeolian research, sample planning should identify the smallest difference between treatments or sampling periods that must be detected to meet project objectives. The level of confidence necessary to make inferences from data should be determined so that sample sizes needed to detect differences in transport can be selected. The scientific and management risks of not using a sample design that can detect change at the level of precision required should also be considered. In all cases, we encourage reporting sample designs, sample sizes, and uncertainty in results so that interpretations are not potentially misleading. If the aim is to monitor net change in erosion, then approaches that use ^{137}Cs or other tracers that are cost-effective for sampling spatial variability in soil properties to detect change should also be considered (e.g., Li et al., 2015).

Finally, it should be acknowledged that it is difficult to establish robust sample designs without knowing the spatial and temporal variances of a property being measured or expected treatment effects. For aeolian transport, variances in Table 2 from diverse land cover types could be used to establish initial sample designs. Future research using the network data will examine the effects of differences in vegetation structure and erodible sediment supply among sites on their spatial variances in aeolian transport. Pilot surveys to measure variances in transport using a large number of samples (e.g., ~30) could also be used when establishing an experiment or monitoring (de Gruijter et al., 2006). Sample designs should optimize the inferential power of data for characterizing aeolian transport rates and detecting change given site characteristics and monitoring goals (Li et al., 2015). However, care should be taken not to oversample to obtain statistically significant effects while ignoring scientific meaning, and sampling should be adapted to ensure that project objectives are met (Lenth, 2001). It should also be recognized that a very high density of sediment samplers could modify site aerodynamics and influence sediment transport rates, making it even harder to measure natural processes accurately. Such effects will be determined by the size and shape (frontal area) of instruments and could be estimated from drag partition theory (e.g., Raupach et al., 1993). Measurement of factors controlling aeolian transport should follow the same sample design principles. Surface aerodynamic roughness, wind friction velocity, threshold friction velocity, vegetation foliar cover and structure all have potentially large spatiotemporal variability (e.g., Gillette, 1999). Implementing sample designs that reduce uncertainty in their measurement would improve confidence in inference about their relations, our understanding of aeolian processes across ecosystems and land uses, and of wind erosion, dune dynamics and dust cycle responses to environmental change.

Acknowledgements

Funding and support were provided by the USDA-ARS LTAR Network, NRCS, BLM, US Geological Survey and Department of Defense. We thank technical staff with the National Wind Erosion Research Network for their efforts in data collection and management and Darren James for discussions about the analysis. Data reported here are available from the National Wind Erosion Research Network Data Portal https://winderosionnetwork.org/data-portal. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

References

Arvin, L.J., Riebe, C.S., Aciego, S.M., Blakowski, M.A. (2017), Global patterns of dust and bedrock nutrient supply to montane systems, *Science Advances*, *3*: eaao1588.

Baas, A.C.W., van den Berg, F. (2018), Large-Scale Particle Velocimetry (LSPV) of aeolian sand transport patterns, *Aeolian Research*, *34*, 1-17.

Belnap, J., Reynolds, R.L., Reheis, M.C., Phillips, S.L., Urban, F.E., Goldstein, H.L. (2009), Sediment losses and gains across a gradient of livestock grazing and plant invasion in a cool, semi-arid grassland, Colorado Plateau, USA, *Aeolian Research*, *1*, 27-43.

Bergametti, G., Gillette, D.A. (2010), Aeolian sediment fluxes measured over various plant/soil complexes in the Chihuahuan desert, *Journal of Geophysical Research*, *115, F03044, doi:10.1029/2009JF001543, 2010*.

Brus, D.J., Noij, I.G.A.M. (2008), Designing sampling schemes for effect monitoring of nutrient leaching from agricultural soils, *European Journal of Soil Science*, *59*, 292-303.

Chappell, A., McTainsh, G.H., Leys, J.F., Strong, C.L. (2003), Using geostatistics to elucidate temporal change in the spatial variation of aeolian sediment transport, *Earth Surface Processes and Landforms*, *28*(6), 567-585.

Chappell, A., McTainsh, G.H., Leys, J.F., Strong, C.L. (2003), Simulations to optimize sampling of aeolian sediment transport in space and time for mapping, *Earth Surface Processes and Landforms*, 28, 1123-1241.

Chappell, A., Baldock, J.A. (2016), Wind erosion reduces soil organic carbon sequestration falsely indicating ineffective management practices, *Aeolian Research*, 22, 107-116.

de Gruijter, J., Brus, D.J., Bierkens, M.F.P., Knotters, M. (2006), *Sampling for Natural Resource Monitoring*, Springer-Verlag, Berlin.

Desu, M.M., Raghavarao, D. (1990), *Sample-size Methodology*, Academic Press, Boston.

Durán, O., Claudin, P., Andreotti, B. (2011), On aeolian transport: Grain-scale interactions, dynamical mechanisms and scaling laws, *Aeolian Research*, *3*, 243-270.

Ellis, J.T., Li, B., Farrell, E.J., Sherman, D.J. (2009), Protocols for characterizing aeolian mass-flux profiles, *Aeolian Research*, *1*, 19-26.

Ellis, J.T., Sherman, D.J., Farrell, E.J., Li, B. (2012), Temporal and spatial variability of aeolian sand transport: Implications for field measurements, *Aeolian Research*, *3*, 379-387.

Gillette, D.A. (1999), A qualitative geophysical explanation for "Hot Spot" dust emitting source regions, *Contributions to Atmospheric Physics*, *72*(1), 67-77.

Gillette, D.A., Herrick, J.E., Herbert, G.A. (2006), Wind characteristics of mesquite streets in the northern Chihuahuan Desert, New Mexico, USA, *Environmental Fluid Mechanics*, *6*, 241-275.

Gonzales, H.B., Ravi, S., Li, J., Sankey, J.B. (2018), Ecohydrological implications of aeolian sediment trapping by sparse vegetation in drylands, *Ecohydrology*, *11*, doi:10.1002/eco.1986.

Goossens, D., Offer, Z.Y., London, G. (2000), Wind tunnel and field calibration of five aeolian sand traps, *Geomorphology*, *35*, 233-252.

Goossens, D., Gross, J., Spann, W. (2001), Aeolian dust dynamics in agricultural land areas in Lower Saxony, Germany, *Earth Surface Processes and Landforms*, *26*, 701-720.

Goossens, D., Buck, B.J. (2012), Can BSNE (Big Spring Number Eight) samplers be used to measure PM10, respirable dust, PM2.5 and PM1.0?, *Aeolian Research*, *5*, 43-49.

Haustein, K., Washington, R., King, J., Wiggs, G., Thomas, D., Menut, L (2015), Testing the performance of state-of-the-art dust emission schemes using DO4Models field data, *Geoscientific Model Development*, *7*, 5739-5789.

Hoffmann, C., Funk, R., Li, Y., Sommer, M. (2008), Effect of grazing on wind driven carbon and nitrogen ratios in the grasslands of Inner Mongolia, *CATENA*, *75*, 182-190.

Hooper, J., Marx, S.K. (2018), A global doubling of dust emissions during the Anthropocene?, *Global and Planetary Change*, *169*, 70-91.

Horvitz, D.G., Thompson, D.J. (1952), A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, *47*, 633-685.

Lenth, R.V. (2001), Some practical guidelines for effective sample size determination, *The American Statistician*, *55*, 187-193.

Kok, J., Ward, D.S., Mahowald, N.M., Evan, A.T. (2018), Global and regional importance of the direct dust-climate feedback, *Nature Communications*, *9*, 241.

Li, Y., Chappell, A., Nyamdavaa, B., Yu, H., Davaasuren, D., Zoljargal, K. (2015). Cost-effective sampling of ^{137}Cs-derived net soil redistribution: part 1 – estimating the spatial mean across scales of variation, *Journal of Environmental Radioactivity*, 141, 97-105.

Martin, R. L., Kok, J.F., Hugenholtz, C.H., Barchyn, T.E., Chamecki, M., Ellis, J.T. (2018), High-frequency measurements of aeolian saltation flux: Field-based methodology and applications, *Aeolian Research*, *30*, 97-114.

Nauman, T.W., Duniway, M.C., Webb, N.P., Belnap, J. (2018), Elevated aeolian sediment transport on the Colorado Plateau, USA: The role of grazing, vehicle disturbance, and increasing aridity, Earth Surface Processes and Landforms, doi:10.1002/eso.4457.

Priestley, M. (1982), Spectral Analysis and Time Series. Academic Press, New York.

Raupach, M.R., Gillette, D.A., Leys, J.F. (1993), The effect of roughness elements on wind erosion threshold, *Journal of Geophysical Research*, *98*, *D2*, 3023-3029.

Sankey, J.B., Germino, M.J., Glenn, N.F. (2012), Dust supply varies with sagebrush microsites and time since burning in experimental erosion events, *Journal of Geophysical Research: Biogeosciences*, *117*, *G01013*, doi:10.1029/2011JG001724.

Sherman, D.J., Li, B., Ellis, J.T., Swann, C. (2018), Intermittent aeolian saltation: A protocol for quantification, *Geographical Review*, *2*, 296-314.

Smeets, C.J. P. P., van den Broeke, M.R. (2008), Temporal and Spatial Variations of the Aerodynamic Roughness Length in the Ablation Zone of the Greenland Ice Sheet, *Boundary-Layer Meteorology*, *128*, 315-338.

Smith, D.R., Snyder, C.D., Hitt, N.P., Geissler, P.H. (2014), Chapter 6: Monitoring for Threshold-Dependent Decisions, in *Application of Threshold Concepts in Natural Resource Decision Making*, edited by G. R. Guntenspergen, pp. 87-100, Springer-Verlag, New York.

Sterk, G., Stein, A. (1997), Mapping wind-blown mass transport by modeling variability in space and time, *Soil Science Society of America Journal*, *61*, 232-239.

Stout, J E., Zobeck, T.M. (1997), Intermittent saltation, *Sedimentology*, *44*, 959-970.

von Holdt, J.R.C., Eckardt, F.D., Baddock, M.C., Wiggs, G.F.S. (2019), Assessing landscape dust emission potential using combined ground-based measurements and remote sensing data, *Journal of Geophysical Research: Earth Surface*, doi:10.1029/2018JF004713.

Webb, N P., Herrick, J.E., Hugenholtz, C.H., Zobeck, T.M., Okin, G.S. (2015), Standard Methods for Wind Erosion Research and Model Development: Protocol for the National Wind Erosion Research Network, USDA-ARS Jornada Experimental Range, Las Cruces, New Mexico.

Webb, N.P., Herrick, J.E., Van Zee, J.W., Courtright, E.M., Hugenholtz, C.H., Zobeck, T.M., et al. (2016), The National Wind Erosion Research Network: Building a standardized long-term data resource for aeolian research, modeling and land management, *Aeolian Research*, *22*, 23-36.

Webb, N.P., Van Zee, J.W., Karl, J.W., Herrick, J.E., Courtright, E.M., Billings, B.J., et al. (2017), Enhancing wind erosion monitoring and assessment for US rangelands, *Rangelands*, *39*, 85-96.

Webb, N.P., Pierre, C. (2018), Quantifying anthropogenic dust emissions, *Earth's Future*, *6*, 286-295.

Webster, R., Oliver, M.A. (1992), Sample adequately to estimate variograms of soil properties, *Journal of Soil Science*, *43*, 177-192.

Woodward, M. (1992), Formulae for sample size, power and minimum detectable relative risk in medical studies, *Journal of the Royal Statistical Society, Series D*, *41*, 185-196.

Yizhaq, H., Ashkenazy, Y., Tsoar, H. (2009), Sand dune dynamics and climate change: A modelling approach, *J. Geophys. Res.*, *114*, F01023, doi:10.1029/2008JF001138.

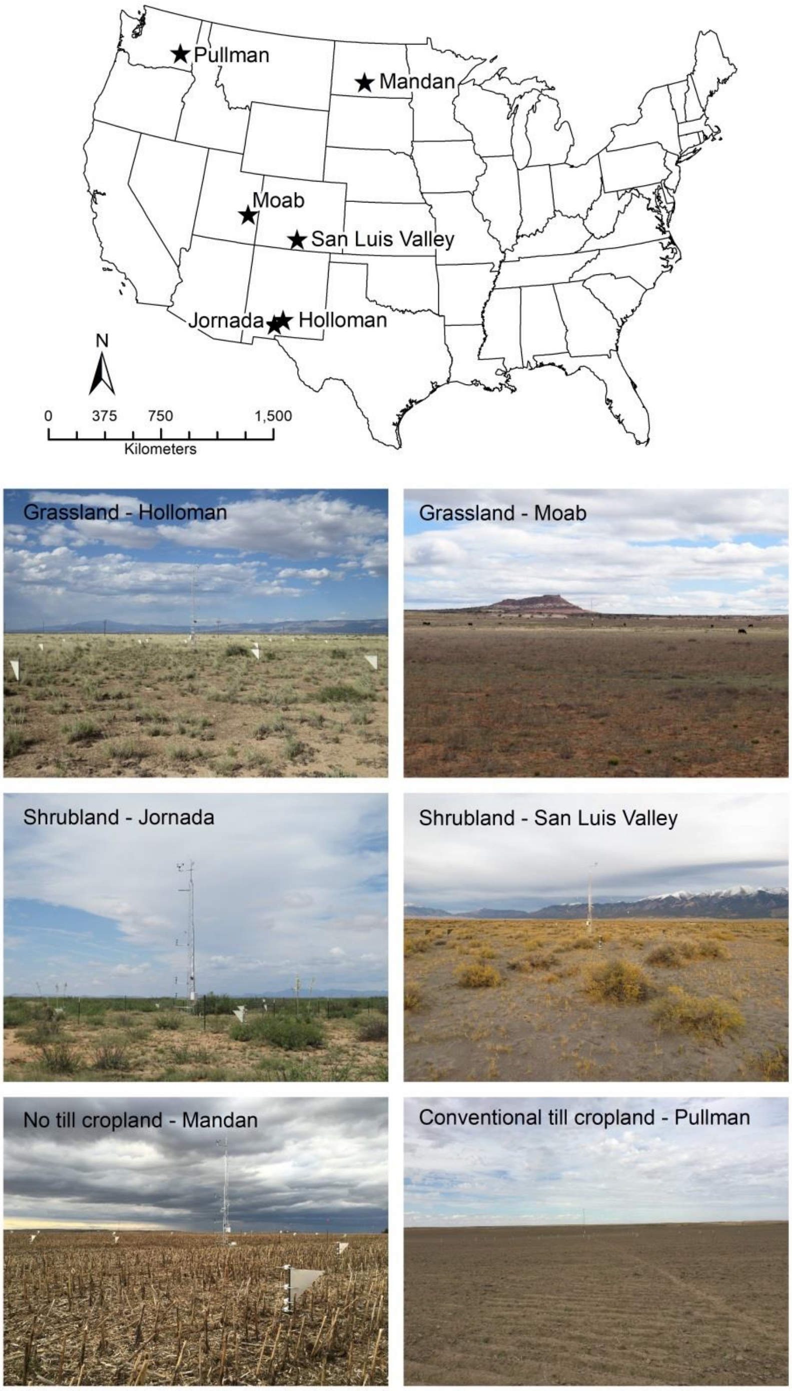

Figure 1 – Locations and photographs of the six US National Wind Erosion Research Network sites from which data were used in this study.

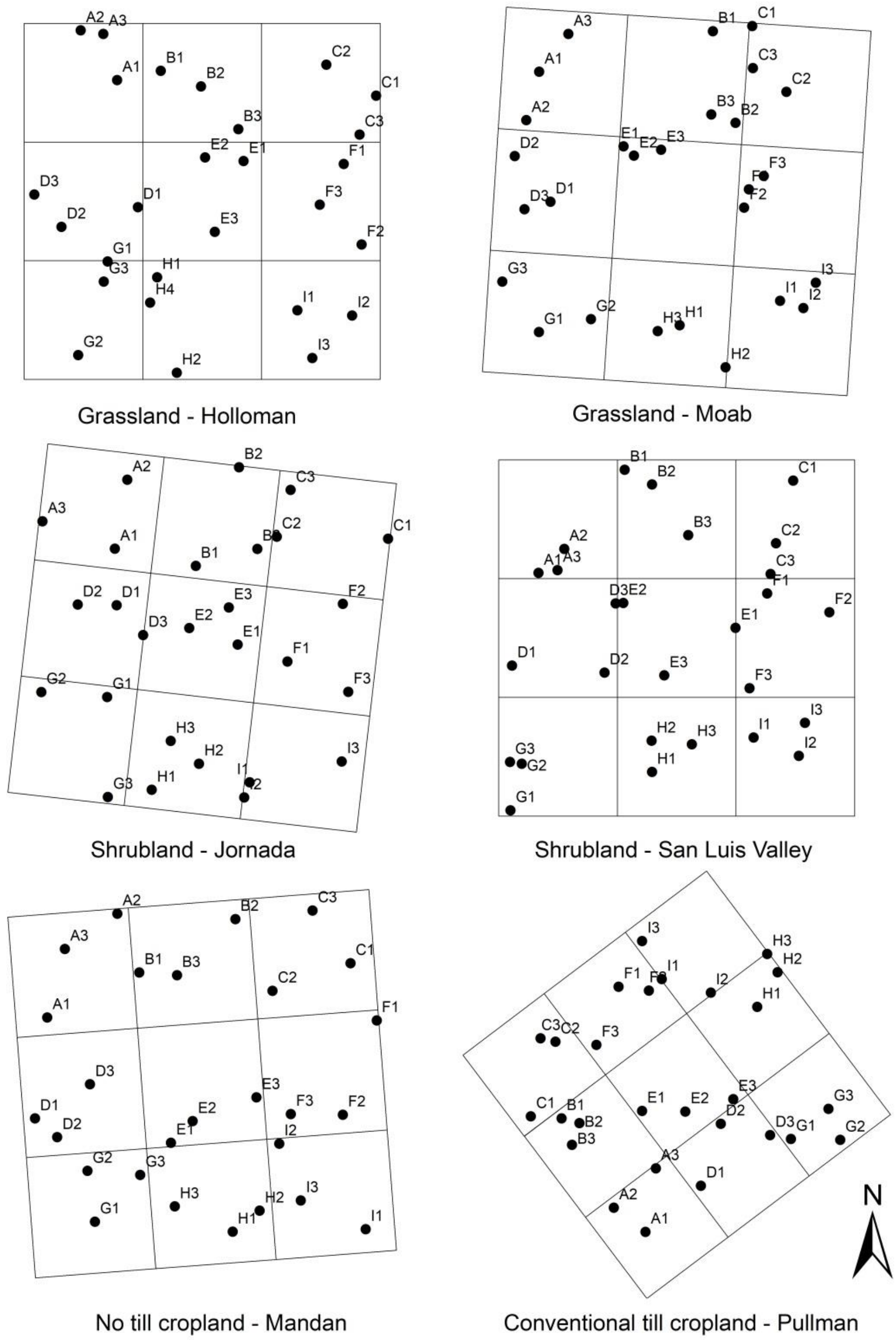

Figure 2 – Schematic showing stratified-random sample design used for Modified Wilson and Cooke (MWAC) sediment sampler masts to measure horizontal sediment mass flux (Q) at the National Wind Erosion Research Network sites. Each site has dimensions of 100 x 100 m.

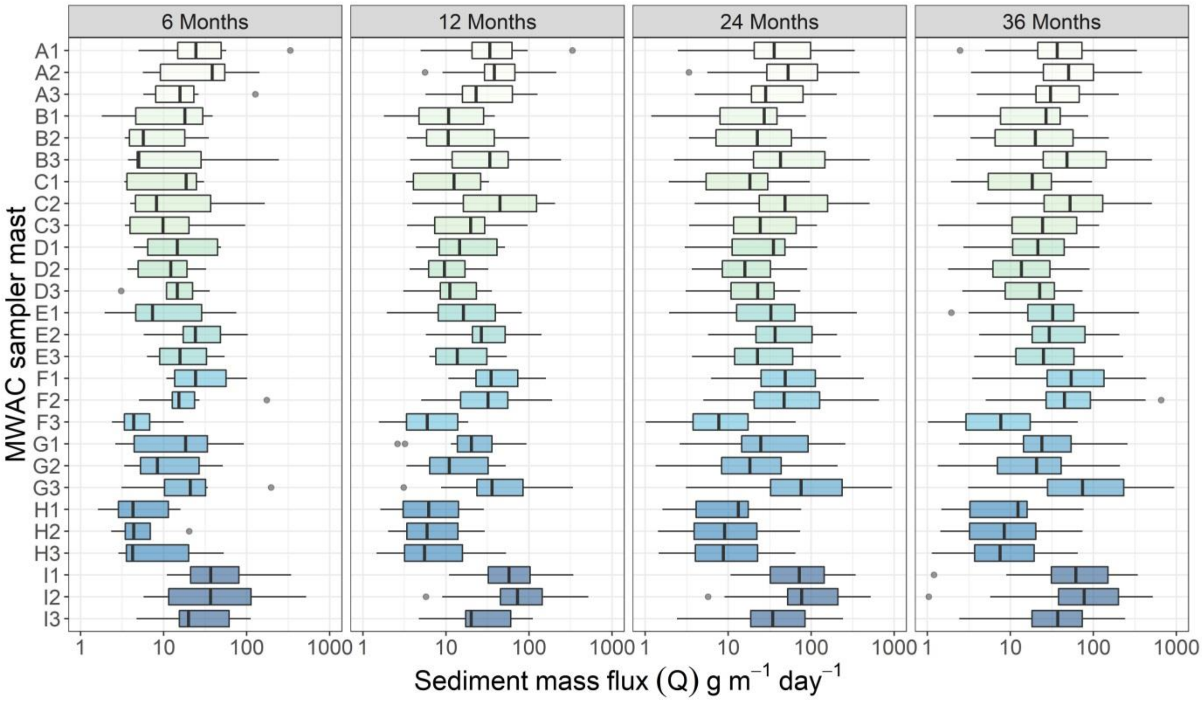

Figure 3 – Boxplots of vertically-integrated horizontal sediment mass flux (*Q*) for each Modified Wilson and Cooke (MWAC) sediment sampler mast at the Jornada shrubland site, summarized over 6 months to 36 months of sampling. Boxes represent interquartile ranges with medians (dark internal lines) and whiskers extend to the range of measurements with dots being data outside 1.5 times the interquartile ranges.

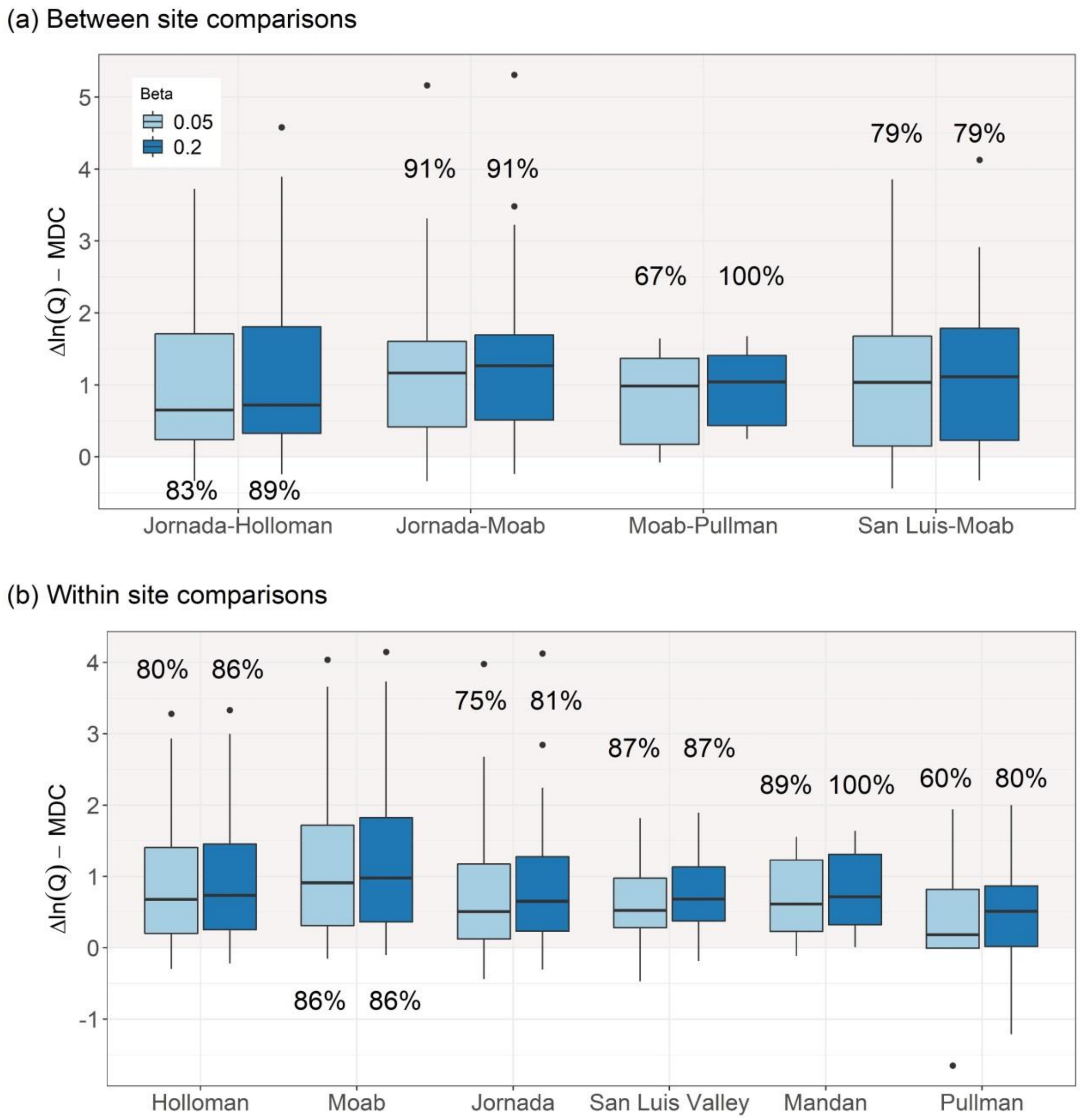

Figure 4 – Distribution of differences between change in horizontal sediment mass flux, $\Delta\ln(Q)$, and the minimum detectable change (MDC) calculated for a one-tailed 5% significance test (Type I error; $\alpha = 0.05$) with 95% confidence and 80% confidence of detecting change between sampling periods (Type II error; $\beta = 0.05$ and 0.2). Results are summarized for (a) between site comparisons and (b) within site comparisons. Positive values indicate that statistically significant differences in aeolian transport were detected. Negative values indicate that statistically significant differences in aeolian transport were not detected. Labels show percentage of comparisons for which statistically significant differences were detected.

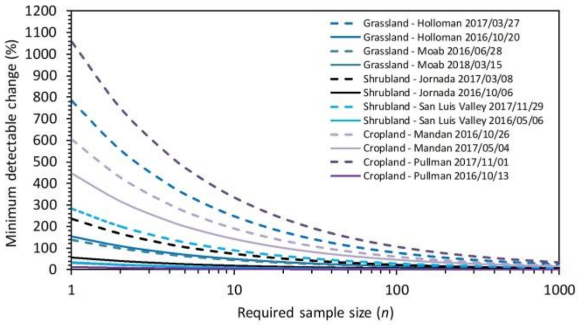

Figure 5 – Response of minimum detectable change (MDC) in horizontal sediment mass flux, ln(Q), for sample size n calculated from small (dashed lines) and large (solid lines) variances measured for a selection of sampling dates at the sites for a stratified random sample with α = 0.05 and β = 0.05. Larger sample sizes enable smaller changes in aeolian sediment transport to be detected with confidence that the changes are statistically significant.

List of Tables

Table 1 – National Wind Erosion Research Network sites and sampling details.

Site name	Holloman (grassland)	Moab (grassland)	Jornada (shrubland)	San Luis Valley (shrubland)	Mandan (cropland)	Pullman (cropland)
Latitude	32.94	38.65	32.63	37.59	46.78	46.89
Longitude	-106.11	-109.87	-106.74	-105.69	-100.95	-118.29
Ecoregion	Chihuahuan Desert	Colorado Plateau	Chihuahuan Desert	Arizona/New Mexico Plateau	Northwestern Great Plains	Columbia Plateau
Management	Rangeland, military land, livestock grazing	Rangeland, livestock grazing	Rangeland, livestock grazing	Rangeland preserve	No tillage cropping	Conventional tillage cropping
USDA soil texture class	Gypsiferous sandy loam	Sandy loam	Sandy loam	Loamy fine sand	Silt loam	Silt loam
Land cover	Grassland	Grassland	Shrubland	Shrubland	Sunflower, wheat, corn	Wheat
Sampling start	08/2015	05/2016	06/2015	06/2016	11/2015	08/2016
Sampling end	05/2018	04/2018	05/2018	11/2017	06/2017	11/2017
Months sampled	36	23	37	16	10	6

Table 2 – Mean ($\bar{Q}$), spatial variances ($\hat{\hat{V}}_s$) and temporal variances ($\hat{\hat{V}}_t$) of vertically-integrated horizontal sediment mass flux (g m^{-1} day^{-1}) for the sites calculated from the stratified random samples. The ratio $\hat{\hat{V}}_t/\hat{\hat{V}}_s$ compares the magnitude of the temporal variances to the spatial variances.

Site	$\bar{Q}$	$\hat{\hat{V}}_s$	$\hat{\hat{V}}_t$	$\hat{\hat{V}}_t/\hat{\hat{V}}_s$
Grassland – Holloman	10.5	10.4	104.9	10.1
Grassland - Moab	21.6	934.4	2,248.0	2.4
Shrubland - Jornada	55.9	4,688.7	7,325.5	1.6
Shrubland - San Luis Valley	179.4	52,018.4	91,324.2	1.8
Cropland - Mandan	7.1	80.3	133.3	1.7
Cropland - Pullman	59.9	1,305.1	4,409.2	3.4